

TESTING THE LIMITS OF ANAPHORIC DISTANCE IN CLASSICAL ARABIC: A CORPUS-BASED STUDY

SAMIR O. JARBOU

Jordan University of Science and Technology
samerjar@just.edu.jo

FATHI MIGDADI

Jordan University of Science and Technology
fhmigdadi@just.edu.jo

Abstract

One of the central aims in research on anaphora is to discover the factors that determine the choice of referential expressions in discourse. Ariel (1988; 2001) offers an Accessibility Scale where referential expressions, including demonstratives, are categorized according to the values of anaphoric (i.e. textual) distance that each of these has in relation to its antecedent. The aim of this paper is to test Ariel's (1988; 1990; 2001) claim that the choice to use proximal or distal anaphors is mainly determined by anaphoric distance. This claim is investigated in relation to singular demonstratives in a corpus of Classical Arabic (CA) prose texts by using word count to measure anaphoric distance. Results indicate that anaphoric distance cannot be taken as a consistent or reliable determinant of how anaphors are used in CA, and so Ariel's claim is not supported by the results of this study. This also indicates that the universality of anaphoric distance, as a criterion of accessibility, is defied.

Key words: anaphora, anaphoric distance, word boundaries, Classical Arabic, Ariel's Accessibility Scale

1. Introduction

There are various types of referring expressions and many of these can be used anaphorically to refer to other structures in discourse. Anaphoric expressions include personal pronouns, noun phrases, and demonstratives. Each of these has its own sub-categories in which each element is associated with specific contexts of occurrence. Most of the research on anaphora attempts to discover the circumstances that necessitate choosing one type of anaphoric expressions, rather than another, in a particular context (see Taboada 2008: 167). In the present study, focus is on one category of referring expressions represented by anaphoric proximal and distal demonstratives; these are investigated in a corpus of CA texts.

In many studies on proximal and distal anaphors, one of the most recurrent notions is that of *accessibility*. In general, the aim of research in this area is to identify the factors that lead a speaker to use particular referential expressions, in connection with specific

antecedent entities, in discourse (see Ariel 2001: 30; and Botley and McEnery 2001: 214). Some of the most detailed analyses of the factors governing the use of referential expressions, including anaphoric demonstratives, are those offered by Ariel (1988; 1990; 2001). The present study tests Ariel's (1988, 1990) claim that anaphoric distance between an anaphor and its antecedent in discourse is a major factor of accessibility and so of determining to use either a proximal or a distal anaphor in context.

In this paper, the determinants of using anaphoric demonstratives in CA will be examined as follows: first, I will explain the major types of anaphoric demonstratives; this is followed by a presentation of Ariel's Accessibility Scale and then a brief description of CA demonstratives. Second, I will provide a description of the CA corpus and details of the method that will be used to measure anaphoric distance for the purposes of this study. Third, I will investigate anaphoric distance in relation to the proximal and distal anaphors used in the CA corpus. This will be followed by a discussion of the results that have been obtained. Finally, I will provide a conclusion and then finish with a brief discussion of the limitations of the present study and suggestions for further research.

2. Background

Demonstratives serve different functions, and so it is important, first, to observe their main types as examined in earlier literature. The general function of demonstratives is "to coordinate the interlocutors' attentional focus in the speech situation" (Diessel 2006: 476; see also Strauss 2002; Diver 1984; Lyons 1978: 636-638; Halliday and Hasan 1976). Interlocutors usually use demonstratives to focus on entities present either in a deictic (i.e. physical) context or a textual (i.e. linguistic) one within discourse. The deictic use of demonstratives is usually referred to as *exophoric* while the textual use is generally called *endophoric* (see Diessel 1999; Himmelmann 1996; Jarbou 2010). It is traditionally known that exophoric demonstrative encode objects or people in the real world; for example, in the utterance "put *this* book on *that* table," *this* and *that* are used to point at objects existing in the physical context of interaction. On the other hand, endophoric demonstratives, which are usually referred to as 'anaphoric', cognitively point at referents (e.g. people, objects, or ideas) associated with antecedent words or expressions in discourse, as will be illustrated in example (1) shortly. Since the present study investigates demonstratives in CA texts, deictic demonstratives are beyond the scope of this study simply because they have "no antecedent that could be measured in the surrounding discourse" (Botley and McEnery 2001: 218). To further define the boundaries of this study, and following the categorization of demonstratives by Dixon (2003), the focus in this paper is on nominal anaphoric demonstratives only; other categories such as adverbial anaphors (e.g. *here* and *there*) are beyond the scope of this study.¹

¹ Dixon (2003) divides demonstratives into (a) nominal demonstratives (b) local adverbials, and (c) verbal demonstratives having the meaning of 'do it like this' (Dixon 2003:62).

The next section provides an overview of the major types of endophoric demonstratives and explains the nature of the relations between these demonstratives and other entities in discourse.

2.1. Anaphors, referents, and antecedents

Endophoric demonstratives are either *anaphoric* or *discourse* demonstratives. These two types of demonstratives are usually described as discourse-referring means “designed to continue the existing attention focus [i.e. center of attention] established hitherto (or assumed to be so established)” in discourse (Cornish 2008: 999). In the case of anaphoric demonstratives in particular, an anaphor (i.e. anaphoric demonstrative) is, mostly, used as a tracking device that refers to “another expression or element in the discourse domain, called the antecedent” (Seuren 2009: 288). The antecedent is present in the “co-text” (Grundy 2000: 27) and anaphors are intended to “continue (sustain) a previously established focus towards” the antecedent mentioned earlier in discourse (Ehlich 1982: 330).² In this relation, an anaphor is a dependent term whereas an antecedent is an autonomous term (Cornish 1996: 21).

The concept of anaphora is generally explained in terms of coreferentiality. That is, a demonstrative is anaphoric when it refers to the same entity that a prior term (i.e. the antecedent) refers to in discourse (Lyons 1978: 660; Levinson 1983:67; see also Levinson 2006; Cornish 2008). In most cases, the anaphor does not refer to the antecedent itself since both mutually point at another entity outside the text, which is the referent. This anaphoric referent, however, has a cognitive nature rather than a physical one since it is “in the universe-of-discourse, which is created by the text” and is not itself located within the text (Lyons 1978: 670).

The relation between anaphors and their antecedents can be *direct* or *indirect*. When the antecedent of an anaphor is easily identifiable (usually when the antecedent is one word), the use of the demonstrative is described as involving direct anaphora, as in the following example:

- (1) A: The man in apartment 222 is a criminal.
B: *That's* a policeman.

In (1), the anaphor *that* is coreferential with the antecedent *the man*. The anaphor is intended to “continue (sustain) a previously established focus towards a specific item” (here *the man*) on which attention had been oriented earlier in discourse (Ehlich 1982:330). The relation between *the man* and *that* in (1) is one of direct anaphora. However, when the antecedent of an anaphor “is more difficult to define directly,” this type of relation is called indirect anaphora where “a reader or hearer may have to carry

² This is similar to the notion of *focus* in Sidner (1983) where *focus* and anaphora “operate to establish and maintain a reciprocal state of shared knowledge about discourse entities in focus between the producer of a text (i.e. a speaker) and the receiver of a text” (cited in Botley and McEnery 2000: 15).

out a somewhat complex process of inference to arrive at the antecedent” (Botley 2006: 74), as in example (2) below:

- (2) A: They say global warming is changing our world.
B: I definitely agree with *that*.

The relation between the anaphor *that* and its referent in (2) is not straightforward or direct. The anaphor in (2) refers to the proposition of an antecedent discourse segment (i.e. the idea that ‘global warming is changing our world’); it does not refer to a clearly defined object represented by the antecedent.³ In (3) below, however, the distal anaphor *that* points at its antecedent itself (i.e. as a discourse segment). That is, it is not coreferential with its antecedent nor does it refer to the proposition expressed by the antecedent.

- (3) A: One of the members said, “the proposal should be double-checked”
B: Who said *that*?

Demonstratives used in the same manner as that in example (3) are typically referred to as *discourse anaphors*. These anaphors are similar to what Lyons (1978) calls “pure textual deixis,” where a demonstrative points at a antecedent discourse segment itself rather than at its referent or proposition. For Levinson (1983), the same linguistic phenomenon represents one category of what he calls “discourse deixis.” Discourse anaphors function meta-linguistically since they point at expressions mentioned in previous discourse (see Botley 2006).

As has been observed in this section, depending on the nature of the relation between an anaphor and its antecedent, endophoric demonstratives are commonly divided into direct, indirect, and discourse demonstratives. In the first two uses, an anaphor is coreferential with its antecedent since these two mutually point at a referent outside the text. A discourse demonstrative, however, is not coreferential with its antecedent since it points at the antecedent itself within discourse. Several theories have been postulated in order to investigate the relation between referring expressions like anaphoric demonstratives and their antecedents in discourse. One of these theoretical accounts, presented in the next section, is that offered by Ariel (1988; 1990; 2001) in her research on referential expressions in English.

2.2. Anaphoric demonstratives within Ariel’s Accessibility Scale

This section presents Ariel’s (1988; 1990; 2001) Accessibility Scale, its categories and sub-categories, and discusses the four factors that Ariel believes determine the use of referential expressions. Ariel’s claim concerning anaphoric proximal and distal demonstratives is also highlighted and discussed.

³ For more information on the different types of direct and indirect anaphora, see Francis (1994); Diessel, (2006); Botley (2006).

Ariel (1988; 1990) draws attention to the notion of *accessibility* to investigate the relation between various referring expressions and their antecedents in a corpus of English. She categorizes referential expressions in an Accessibility Scale according to their degree of retrievability.⁴ *Retrievability* generally refers to the degree of cognitive effort, or difficulty, by which an antecedent may be remembered or recalled by the addressee or reader. Ariel (1988; 1990; 2001) believes that some antecedents “will require less processing effort to retrieve than others, and this will influence which referring expressions are used” (Botley and McEnery 2001: 214). According to Ariel (1988; 1990), there are four determinants of retrievability or accessibility of referential expressions: the first is *anaphoric* (or *textual*) *distance* which is represented by the number of words and clauses between an anaphor and its antecedent, the second factor is *competition* between the different possible antecedents, the third factor is *saliency* which refers to whether the antecedent is topic or non-topic, and the fourth factor is *unity* which is about whether the antecedent and its anaphor are within the same world, point of view, segment, or paragraph.

Although Ariel (1988: 65; 2001: 33-38) mentions that accessibility depends on more than one factor, she mainly focuses on anaphoric distance. Based primarily on anaphoric distance, Ariel (1990: 73; 2001: 31) categorizes referring expressions into an Accessibility Scale where pronouns are High Accessibility Markers, anaphoric demonstratives are Mid Accessibility Markers, and proper names and definite descriptions are Low Accessibility Markers. The easier it is to retrieve the antecedent of a referring expression, the higher is the degree of its accessibility within Ariel’s Scale. For instance, it is easier to retrieve or recognize the antecedent of a third-person pronoun (which is a High Accessibility Marker) such as *they* in “the little boys came early and *they* wanted to go to the beach” than to retrieve the antecedent of the distal demonstrative (which is an Intermediate Accessibility Marker) in “he said the crops will be ruined by rain, but nobody believed *that*”. These uses of *they* and *that* can be compared with the higher difficulty in retrieving the antecedent of a definite description (i.e. Low Accessibility Marker) such as *the teacher* when its antecedent occurs in a previous paragraph with many intervening sentences between this definite description and its antecedent (for more examples, see Ariel 1990: 7-8).

⁴ A similar representation is Gundel et al.’s (1993) *Givenness Hierarchy* which aims to illustrate the criteria by which a referring expression is chosen to refer to a particular entity in discourse based on the cognitive status of the referent. According to Gundel et al. (1993), using referring expressions reflects different cognitive statuses; these are arranged to form a Givenness Hierarchy consisting of six possible cognitive statuses that assist the addressee in recognizing the antecedent in discourse. However, unlike Ariel’s Accessibility Scale, “the difficulty with Gundel’s work is in finding testable claims that can be evaluated in a corpus. Unlike Ariel, Gundel does not give any specific metrics, such as textual distance, to allow us to measure the extent to which particular anaphoric expressions reflect particular cognitive statuses” (Botley and McEnery 2000: 10). Another area of study that can be related to Ariel’s Accessibility theory and Gundel et al.’s work is Centering Theory which is “a theory of local focus in discourse that proposes different transition types between any pair of utterances. Those transitions are based on salience, but also on the expectations that the hearer might have about the focus of the next utterance” (Taboada 2008: 177); see also Grosz et al. 1995; Walker et al. 1998; and Taboada et al. 2008.

Ariel believes that the criteria she uses to construct the Accessibility Scale (i.e. as a macro structure) also apply to the categories within the Scale in that each of these has its own internal scale (i.e. a micro-scale) as well. In other words, each of the individual categories has its own sub-categories of forms that are also ordered according to their accessibility values. For instance, anaphoric demonstratives, representing Mid Accessibility Markers, are “themselves distinguished as to Accessibility” since “proximal and distal Markers serve to refer to entities over shorter and longer textual distance respectively” (Ariel 1988: 76). Based on Ariel’s (1988; 1990; 2001) Accessibility Scale, proximal demonstratives are High Accessibility Markers while distal demonstratives are Low Accessibility Markers within the main category of Mid Accessibility Markers. She uses anaphoric distance as the determining factor to categorize referential expressions into High, Intermediate, and Low accessibility Markers. The same criterion (i.e. distance) is also used to label the sub-categories within each of these three categories in Ariel’s Scale.

Consequently, as far as the accessibility value of anaphoric demonstratives is concerned, Ariel (1988; 1990; 2001) focuses mainly on anaphoric distance; she indicates that the other three factors (i.e. *saliency*, *competition*, and *unity*) are far more related to either High Accessibility or Low Accessibility Markers than to Intermediate Accessibility Markers.⁵ Ariel constructs the internal taxonomy of Intermediate Accessibility Markers (i.e. demonstratives) primarily by depending on the factor of anaphoric distance (see Ariel 1990: 31). According to Ariel’s proposal, within this internal taxonomy of demonstratives, the factors of *saliency*, *competition*, and *unity* are far less relevant than *anaphoric distance* in determining the accessibility status of proximals and distals (see Ariel 1990: 69-72). In the present study, the factors of *competition* and *saliency* are naturally neutralized since they concern cases where an anaphor can possibly refer to more than one structure that can compete for the position of an antecedent. These cases, however, as Ariel indicates, concern, and call for the use of, High Accessibility Markers rather than Intermediate ones such as demonstratives (Ariel 1990: 69-72).⁶

On the other hand, Ariel (1990) indicates that the factor of *unity* applies more to Low Accessibility Markers such as full names (i.e. first name followed by family name) than to Intermediate or High Accessibility Markers. *Unity* concerns cognitive statuses of memory where the speaker, for one reason or another, perceives that it is easy (i.e. in cases of strong unity) or, conversely, it is difficult (i.e. in cases of weak unity) to retrieve the antecedent. In cases of low accessibility, the antecedent is sometimes in a paragraph previous to the one where the anaphor is, and also there are usually many intervening sentences between the antecedent and its anaphor. This long anaphoric distance would

⁵ Ariel uses *saliency* in addition to *anaphoric distance* to label the sub-categories within the internal taxonomy of High Accessibility Markers. She also uses *unity* in addition to *distance* to label the sub-categories within Low Accessibility Markers.

⁶ Moreover, even in cases where there could have been competitive antecedents for any of the CA demonstratives and a decision had to be made as regards their relevant degree of saliency, these had not been included in this study. These occurrences have been very rare in the corpus as far as anaphoric demonstratives are concerned and they have been excluded from data analysis since any attempt to recognize the definite antecedents of these demonstratives always led to ambiguity and vagueness (see section 3.2 below).

adversely influence unity and accessibility thus suggesting that less unity leads to the need to use Low Accessibility Markers to guarantee retrievability. The factor of *unity* in Ariel is also mostly related to anaphoric distance since long anaphoric distance usually leads to less unity and vice versa in the case of Low Accessibility Markers. As such, unity, investigated in relation to anaphors and antecedents separated by long anaphoric distance, cannot be applied to the study of the accessibility degree of Intermediate Accessibility Markers.⁷ Ariel (1988; 1990) does, however, mention that there are other criteria to measure unity between referential expressions and their antecedents but she does not investigate these in relation to demonstratives.

The present study aims to test Ariel's (1988; 1990) claim concerning anaphoric distance by applying it to an analysis of the use of anaphoric demonstratives in CA. Ariel's claims can be summed up as follows: anaphoric distance is a significant determinant of accessibility since proximal anaphors tend to be associated with antecedents at short distance while distals tend to be associated with antecedents at long distance in discourse.

In order to test the applicability of this claim with regard to CA anaphors, this study aims to answer the following basic question: are anaphoric demonstratives in CA used in relation to anaphoric distance? That is, is anaphoric distance a reliable criterion to explain differences in using proximal and distal anaphors in CA?⁸ The hypothesis in this paper is that, unlike the predominant tendency which Ariel (1988; 1990) discovers for anaphors in English, anaphoric distance cannot be taken as a determinant for the use of proximal and distal anaphors in CA. Sources of data for this study of CA anaphors are represented by a huge corpus of prose texts (see section 3 below).

To sum up, Ariel (1988; 1990) claims that anaphoric distance is a major determinant concerning the use of anaphoric demonstratives in discourse, as proximals are associated with antecedents at short distance while distals are associated with antecedents at long distance. Anaphoric distance is represented by the number of words and clauses between an anaphor and its antecedent. Based mainly on anaphoric distance values, Ariel labels proximals as High Accessibility Markers while distals are Low Accessibility Markers. Ariel's claim will be examined (see section 4) in relation to how demonstratives are used in a corpus of CA texts to test the reliability of anaphoric distance as a determinant of the use of anaphors in CA. Before doing this, the next section provides an overview of demonstratives in CA.

2.3. Demonstratives in CA

Arabic belongs to the family of West Semitic languages which also includes other languages like Aramaic and Hebrew. CA was the language of poetry, public speeches,

⁷ See also section (4) below for findings concerning the effect of the distance factor on the accessibility degree of demonstratives in CA.

⁸ This study is not concerned with investigating the overall status of demonstratives themselves as Intermediate Accessibility Markers in comparison with other categories of referential expressions; focus is on the elements within the internal (i.e. proximal/distal) taxonomy of this category.

and literary productions in general (see Owens 2006: 38) in the Arabian Peninsula.⁹ This variety was used before and shortly after the rise of Islam in the seventh century AD; the generally “accepted hypothesis is that it represents a sort of *koine* or common language used for poetry (and probably for ‘elevated diction’ in general) throughout the peninsula” (Beeston 1970: 13; see also Brockelmann 1977: 42, vol. I). This suggests that CA was a common and widespread literary variety and that its users also had their own dialectal or everyday varieties of spoken Arabic (see Beeston 1970: 13; Brockelmann 1977: 42, vol. I).

The number of nominal demonstratives in CA is actually large since it has more than 30 demonstratives (see Appendix B).¹⁰ Many of these have rare occurrences in CA; their demonstrative functions are usually taken over by the following far more common demonstratives: *haaḍa* ‘this-Sg. M.’, *haaḍihi* ‘this-Sg. F.’, *haaḍaan*, ‘these-Dl. M.’, *haataan* ‘these-Dl. F.’, *ḍaalika* ‘that-Sg. M.’, *tilka* ‘that-Sg. F.’, *haaʔulaaʔi* ‘these-Pl. M./F.’, and *ʔulaaʔika* ‘those-Pl. M./F.’¹¹ These eight demonstratives are attested in the corpus. However, the focus in this paper is only on the following four singular demonstratives: *haaḍa*, *haaḍihi*, *ḍaalika*, and *tilka*. The other four (i.e. dual and plural) anaphoric demonstratives have not been taken into consideration in this study for two main reasons: first, they are scarce in the corpus, and, second, they are expected to exhibit the same behavior as that of singular anaphors.

Many of the demonstratives in CA share the same lexical bases to which different types of affixes are attached. Some of these affixes represent inflections for gender and number. Dual demonstratives, unlike singular and plural ones, are also inflected for case. Many of the demonstratives in CA are formed by adding affixes to the proximal bases *ḍa* ‘this-Sg. M.’, *ḍee* ‘this-Sg. F.’ and *ʔul* ‘these-Pl. F./M.’ (for details see Jarbou 2012). This behavior of CA demonstratives reflects the general behavior within West Semitic languages (see Hasselbach 2007: 24). Other affixes that are usually attached to proximal bases in CA are *haa-*, *-li-*, *-ka*, and *-n-*. The prefix *haa-* is optionally added to proximals, does not have any distance-indicating value, and functions as an attention-getter. The suffixes *-li-* and *-ka* are added to singular and plural demonstratives while *-n-* is sometimes infix to dual demonstratives. These three morphemes are distance-indicating (see Jarbou 2012).¹²

⁹ This is a broad description of CA. It is beyond the scope of this paper to provide a definition of CA or to study its historical origins and development. For information on the history of CA see Versteegh 2001; Owens 2006.

¹⁰ A good number of demonstratives in CA had the same semantic function but differed in form since many of these represent dialectal variations in Old Arabic (e.g. Old Hijaazi and Old Najdi) rather than stand for different functions. For instance, the following eight different forms of demonstratives had the same function which is pointing at a proximal, singular, feminine entity: *ḍih*, *ḍii*, *tihi*, *ḍihi*, *tii*, *ḍaat*, *tih*, *taa*.

¹¹ In the Arabic transliteration, vowels are represented as either short or long as follows: /a/-/aa/, /i/-/ii/, and /u/-/uu/. In addition, diphthongs are transliterated as /aj/ and /aw/. See Appendix A for the description of phonemic symbols and abbreviations used in this paper.

¹² A detailed analysis of the historical origin or morphological developments related to demonstratives in CA or Old Arabic varieties is beyond the scope of this paper.

After this short overview of demonstratives in CA, in the next section I turn to describing the details of the method that has been followed to examine the effect of anaphoric distance on using demonstratives in CA.

3. Method

The prose corpus considered for analysis in this study is part of what Owens (2006: 38) refers to as “the early *ʔaadab* or belles lettres literature” of CA. This corpus consists of texts that are included in the first volume of a huge, three-volume collection referred to as *Jamaharat Khotab Al-Arab (JKA)* (literally ‘the multitude of the orations/sermons of the Arabs’). *JKA* is the largest and most comprehensive collection of CA prose. The productions or texts in *JKA* come from different resources such as writings of grammarians of Arabic, philosophers, historians, and from books describing the different aspects of the Arab and Islamic life. A. Safwat compiled these texts and published them in 1933. Each of the volumes of *JKA* deals with a specific historical period. The texts in *JKA* represent real speeches, orations, and sermons (by kings, tribal leaders, and other prominent figures in society) in addition to narrations of prominent events, debates, arguments, descriptions, etc. that had originated in the period from the sixth century A.D. to the ninth century A.D.

The history of the examined texts dates from the pre-Islamic period to the period shortly after the rise of Islam in the Arabian Peninsula. This period covers a time span that spreads from the sixth century A.D. to the end of the seventh century. Most of the texts included in the first volume had been produced within the two major regions in the Arabian Peninsula: Hijaaz (western region of present-day Saudi Arabia) and Najd (central and eastern regions of present-day Saudi Arabia). The examined corpus represents the first half of texts in the first volume. The reason why focus was only on these texts is, firstly, because these texts are taken as representing (early) CA in the peak of its development and before language changes started to take place in the centuries that followed and, secondly, this corpus of about 40,000 words is believed to be large enough to allow for an adequate investigation of how anaphoric demonstratives had been used in CA. The corpus is in the form of texts on paper. It has been analyzed manually simply for the lack of any software that can locate antecedents of anaphoric demonstratives in Arabic, not to mention locating antecedents of indirect or discourse anaphora.

The method of data analysis followed in this paper consists of two steps: taking Ariel’s investigation of accessibility that focuses mainly on anaphoric distance as a model, the first step is to measure the anaphoric distance between the CA anaphors and their antecedents as they occur in the corpus. How this was actually performed will be discussed in the next section and the results will be presented in section 4.

3.1. Measuring anaphoric distance

The issue of how to measure anaphoric distance between referring expressions and their antecedents is relatively debatable (see Taboada et al. 2008). For instance, in Givón (1983), anaphoric distance is measured based on the number of clauses between

anaphoric noun phrases and their antecedents. In another study, Botley and McEnery (2001) calculate anaphoric distance based on the number of words and sentences between anaphoric demonstratives and their antecedents.

In the present study, which focuses on demonstratives rather than on a variety of expressions that includes other types of referring expressions, anaphoric distance was calculated by counting words, rather than phrases, clauses, or sentences, between anaphors and their antecedents. It is assumed here that sentence or clause count would not give accurate measurements of anaphoric distance between anaphoric demonstratives and their antecedents, at least in CA, simply because it is common for both proximals and distals to have their antecedents in an immediately preceding sentence or clause. In other words, there are frequent occurrences, in the corpus, of proximal and distal anaphors that have the same or similar sentence-distance (see examples 6 and 7 below). For that reason, their anaphoric distance could only be measured through word-distance. Word count is here perceived as being far more “computationally feasible” (see Botley and McEnery 2001: 232) than counting sentences or clauses to measure anaphoric distance between demonstratives and their antecedents.¹³

Moreover, it is expected that sentence, clause, or phrase-count would work better when studying the relative accessibility statuses of anaphoric expressions within the category of Low Accessibility Markers. Within this latter category, an antecedent and its anaphor are usually separated by (many) intervening phrases and sentences and sometimes they are also separated across paragraph boundaries. The next section provides the definition of *word* that has been utilized in this study for the purposes of counting words and recognizing word boundaries in the CA texts.

3.2. Words and word boundaries in CA

In any single study that depends on word count to measure anaphoric distance, one needs to decide on what constitutes a word boundary. This is important in order to have a reliable tool that gives consistent measurements of anaphoric distance. However, to recognize word boundaries in a corpus of written texts, it is essential to have a practical and effective definition of what a *word* is. The definition of what constitutes a word, which has been adopted for the purposes of this study, combines the features of morphological and orthographic words.

An orthographic word is represented by a written sequence of letters which is “preceded by a blank space and followed either by a blank space or a punctuation mark” (Plag 2003: 4). In morphology, however, a morpheme (i.e. the smallest meaningful sound unit) is divided into two types: free morphemes (e.g. ‘nature’), that can stand independently in speech, and bound morphemes (mostly affixes) (see Plag 2003:10). Accordingly, a morphological word can be described as a free morpheme which has or does not have bound morphemes attached to it.

In addition, as far as the grammarians of Arabic are concerned, a word or *kalima* in Arabic is *ʔima ʔisim wa ʔima fiʔil wa ʔima harf* ‘*duu maʔna*’; this can be translated as

¹³ However, as mentioned earlier in this section, Botley and McEnery (2001) use word and sentence count.

follows: a word in Arabic is ‘either a noun, a verb, or a letter (i.e. particle) that has meaning’ (see Ibn 3aqiil 1996: 13-16). This definition identifies words as independent syntactic categories that have meaning. These categories are represented by parts of speech. This definition also indicates that a word in Arabic consists of one free morpheme that might have other morphemes attached to it. It should be taken into account that verbs in Arabic are inflected for tense, person, number, and gender. These inflections are affixed to the root of the verb to form one morphological word. For instance, although personal pronouns in Arabic are sometimes attached to verbs, any of these verbs would still be considered as one morphological, or orthographic, word. For example, *ʔkalaha* ‘he ate it’ is transliterated as ‘ate-he-it’ but is considered as a single morphological or orthographic word in Arabic.¹⁴ Another example is that the word *biraʔsihi* ‘with his head’ (transliterated as ‘with-head-his’) is one word in spoken and written Arabic since it consists of the free morpheme *raʔs* ‘head’ which is a noun prefixed with the preposition *bi* ‘with’ and suffixed with the pronoun *hi* ‘his’. In view of that, a *word* in Arabic is a morphological word which is represented in writing by a space before it and a space after it.

All parts of speech in an utterance or sentence in Arabic are morphological words, provided that they are not attached to other parts of speech. In the present study, the definition of a *word* that is adopted for the purposes of defining word boundary, which in turn is necessary for measuring anaphoric distance, is as follows: a word in an Arabic utterance or (written) sentence consists of one free morpheme that might have affixes attached to it. Based on common features between the orthographic and morphological accounts of what a *word* in CA is, and for the purposes of this paper, an orthographic word is the same as a morphological word. Examples on these morphological words in CA include content words like nouns, adjectives, and verbs in addition to function words like particles and prepositions (i.e. as independent morphemes). Throughout the whole process of data analysis in this study, anaphoric distance has been measured consistently by counting all orthographic words between each anaphor and its antecedent. The average anaphoric distance of proximals has been compared with that of distals.

However, it is unlikely that all researchers studying anaphoric expressions in different human languages could have followed the same criteria to measure anaphoric distance; some would depend on word count, others on sentence count, and still others would depend on both sentence and word count. Researchers concerned with anaphoric distance do not seem to put much emphasis on whether it is word or sentence count that is used to measure anaphoric distance. What matters most is that the criteria used to measure anaphoric distance for one type of referential expressions (e.g. proximals) be the same that is used for another type (e.g. distals) in the same study so that one can finally have comparable results for the investigated types of expressions. In any study on anaphoric distance, accuracy in comparing and classifying types of anaphoric structures depends on consistency in applying the same measurement tool to the items to be compared (see, as cited in Ariel (1990), Hinds (1983) on Japanese, Bentivoglio (1983)

¹⁴ However, personal pronouns in Arabic can also be independent morphemes. This happens mostly in cases when they, as subjects, occur before their verbs. In these contexts, native speakers recognize the personal pronoun as one morphological word that is also represented as a single orthographic word in writing.

on Spanish, Jaggar (1983) on Hausa, and Gasser (1983) on Amharic, in addition to Ariel (1988, 1990) on English).

The texts included below in examples (4) to (7) can be taken as illustrations on how anaphoric distance has been measured using word count.¹⁵ For instance, consider the following text (from Safwat 1933: 106) on the distal anaphor *ḍaalika*:¹⁶

- (4) *thumma ina 3amran dakhala Hadiiqatan lah wa*
 then be **3amr entered-he garden his and**
ma3ahu ḍgaariyatayni min ḍgawaariih fabalagha ḍaalika
with-him maids-two from maids-his so-reached that
Tariifa
Tariifa
 ‘3amr then entered one of his gardens with two of his maids, and so Tariifa knew of that’

In example (4), the distance between the distal anaphor *ḍaalika* ‘that’ and its antecedent is one orthographic word only represented by *fabalagha* ‘so-reached’. Another example illustrating how anaphoric distance has been calculated according to the definition of *word* and *word boundary* followed in this study is the following text (from Safwat, 1933: 69):

- (5) *ʔayuaridḡaali ʔHabu ʔilajki Qaalat ʔsahlu ʔnnajiib ʔsamihu alHasiib*
 which men prefer to-you said-she **amenable clever meek thoughtful**
ʔnnabu alʔariib Qaala laha hal baQya ʔHadun ʔafdhalu min haaḍa
principled witty said-he to her is remained anyone better than **this**
 ‘what type of men do you think is best? She answered ‘the one who is amenable, clever, meek, thoughtful, principled, and witty’. Then he asked ‘is there anyone who is better than this?’

In (5), the anaphoric distance between the last word in the antecedent *ʔsahlu ... alʔariib* ‘easy-going... witty’ and the proximal anaphor *haaḍa* ‘this’ is five words represented by *hal baQya ʔHadun ʔafdhalu min* transliterated as ‘is remained anyone better than’. In this sequence of five words, we can identify five free morphemes of which two (i.e. *hal* and *min*) are function words and three (i.e. *baQya*, *ʔHadun*, and *ʔafdhalu*) are content words.

In cases of indirect and discourse anaphora, anaphoric distance was measured by counting words between the anaphor and the last word in the phrase or clause standing for the antecedent. Nevertheless, this was sometimes difficult since, in some cases of

¹⁵ Interlinear word-by-word gloss (i.e. transliteration) immediately follows the phonemic transcription of CA examples. Morphemes affixed to the roots of CA words are transliterated, rather than represented by symbols, after a hyphen in the gloss for purposes of clarity since the purpose of providing the gloss is mainly to know the exact number of words that exist between an anaphor and its antecedent. The hyphen indicates morphemic boundary. Morphemes that are joined by a hyphen represent one orthographic word.

¹⁶ For the purposes of illustrating how anaphoric (i.e. word) distance has been calculated, anaphors and their antecedents appear in bold in examples (4) to (7) in this paper. In some cases, this does not mean that all words between the antecedent and anaphor should be included in word count since some of these words represent the narrators’ words that are intended to report what interactants say. However, the narrator’s words are easily recognizable in the transliterations.

indirect anaphora, the boundaries of the antecedent were indistinct and ambiguous. These cases, therefore, were not included in the analyzed data. They appear in Table (1) below underneath the category labeled as ‘unresolved’. Botley (2006: 102) reports similar results of “fuzziness” and “ambiguity” concerning the study of indirect anaphora in English since he believes that “indirect anaphora definitely poses difficulties for corpus-based linguistics, in that almost 30% of IA [i.e. Indirect Anaphora] cases analyzed were hard to classify straightforwardly.”

Anaphor	Number of occurrences	Anaphoric use	Deictic use	Unresolved
<i>haaḍa</i>	142	56	72	14
<i>ḍaalika</i>	181	122	11	48
<i>haaḍihi</i>	33	22	8	3
<i>tilka</i>	23	16	2	5
Total	379	216	93	70

Table 1: Distribution of anaphoric and deictic CA demonstratives in the investigated corpus.¹⁷

The antecedents categorized as ‘unresolved’ in Table (1) were hard to define for two reasons: first, the existence of more than one linguistic constituent that can be taken as representing the antecedent; second, it was difficult sometimes to define the structural boundaries (i.e. beginning and end) of the antecedent. This step of excluding such occurrences of demonstratives, where the antecedent co-exists with other structures that can compete for the position of a more or less salient antecedent status, has further aided in neutralizing the possible effects of *saliency* or *competition* in determining accessibility.¹⁸

To sum up, this section has defined, and provided illustrations on, how word count has been used as a tool to measure anaphoric distance in the CA corpus. In section 4 below, proximal and distal anaphors in the CA corpus are investigated in relation to Ariel’s claim concerning the effect of anaphoric distance on the choice and use of anaphors in discourse.

¹⁷ It is noticeable in Table (1) that the number of occurrences of the feminine demonstratives in the corpus is much less than that of the masculine ones. Although searching for explanations is beyond the scope of this paper, it seems that one reason is that texts in the corpus are mainly about the public and ‘political’ everyday life of Arab tribes that were dominated by men; these texts are mainly about issues related to the lives of, and activities mostly performed by, men such as battles, wars, travels, etc. It can be assumed that women did not have any influential or prominent presence in the public life in Old Arab societies. Consequently, women, and demonstratives referring to them, have a limited presence in those CA texts.

¹⁸ These two factors (i.e. saliency and competition) as explained in 2.2 above, are, expectedly, irrelevant as determinants for the choice between proximals and distals within the class of demonstratives as Intermediate Accessibility Markers.

4. Results and discussion

This section examines Ariel's (1988; 1990) claim (see section 2.2 above) by investigating anaphoric distance related to the use of the CA anaphors *haaḏa*, *ḏaalika*, *haaḏihi*, *tilka*. To achieve this goal, the following method has been performed in this study: after anaphoric distance was measured for each particular instance of proximal and distal anaphors, it was added to the other distance measurements for the same anaphor (i.e. either proximal or distal) in all contexts of its occurrence in the corpus. Then, the average anaphoric distance for each anaphor was calculated by dividing the total sum of distance measurements on the number of instances of that anaphor. For example, the number of occurrences of the anaphor *ḏaalika* 'that' is 122 and the sum total for distance measurements for all of these is 422 words. Consequently, the average textual distance for the anaphor *ḏaalika* is 3.4 words. The same has been done for the other three anaphors. After that, the average anaphoric distance measurements related to proximals were compared with those related to distals to know if this indicated any correlations between anaphoric distance and choice of anaphoric demonstratives.

The average anaphoric distance for each of the investigated anaphors is shown in Table (2) below. The average anaphoric distance for the proximal masculine *haaḏa* 'this-Sg. M.' and for its distal counterpart *ḏaalika* 'that-Sg. M.' is, respectively, 4.1 and 3.4 words. On the other hand, the average anaphoric distance for the proximal feminine *haaḏihi* 'this-Sg. F' and for its distal counterpart *tilka* 'that-Sg. F' is 3.1 and 4.2 words, respectively. Anaphoric distance measurements for the masculine pair show that average distance between the proximal anaphor *haaḏa* and its antecedents is longer than that between the distal anaphor *ḏaalika* and its antecedents. However, the feminine anaphors show the opposite, as the average distance between the proximal anaphor *haaḏihi* and its antecedents is shorter than that between the distal anaphor *tilka* and its antecedents in discourse.

Anaphor	No. of occurrences	Total word distance	Average anaphoric distance (in words)
<i>haaḏa</i> 'this'	56	230	4.1
<i>ḏaalika</i> 'that'	122	422	3.4
<i>haaḏihi</i> 'this'	22	69	3.1
<i>tilka</i> 'that'	16	68	4.2

Table 2: Average anaphoric distance measurements for the proximal and distal anaphors in the CA corpus.

The results of average distance measurements between the proximal and distal masculine anaphors and their relevant antecedents in the CA corpus are similar to those obtained by Botley and McEnery (2001) in their study of anaphoric demonstratives in English. They conclude that "on the whole, the average sentence and word distances for proximal forms are greater than those for distal forms, which runs counter to Ariel's first claim" (Botley and McEnery 2001: 223).

On the other hand, results of the anaphoric distance between the CA feminine anaphors and their antecedents support Ariel's (1988) finding that distal anaphors tend to

refer more to textually remote antecedents while proximal anaphors tend to refer more to textually close antecedents. Nevertheless, the fact that the anaphoric distance measurements for the CA feminine anaphors contradict those for the masculine anaphors shows that anaphoric distance cannot be taken as a consistent, adequate, or comprehensive determinant of accessibility concerning CA anaphors and their antecedents. Another related aspect of this inconsistency of anaphoric distance as a measure of accessibility concerns the positions of anaphors in relation to their antecedents. To be more specific, proximal and distal anaphors in CA sometimes occur in contexts of similar anaphoric distance, as in, below, examples (6) and (7) (from Safwat 1933: 41 and 106, respectively):

(6) *faQaala* *ʔima* *ʔiHiaaʔi* *kulayb* *fahaaða* *maa* *laa* *yakuun*
 said-he as-for resurrecting-I Kulaib so-this never not be
 ‘he said, ‘as for resurrecting Kulaib, this is something that cannot be done’

(7) *madhat* *ʔilaa* *ʔan* *dakhalat* *ʔala* *ʔamru* *wa*
 walked-she to when entered-she on ʔamru and
ðaalika *Hiin* *ʔintaSafa* *ʔanahaar*
 that when at- middle the-day
 ‘she walked away until she (reached and) entered ʔamru’s lodging and that was at midday’

The anaphoric distance between the proximal *haaða* and its antecedent in (6) and that between the distal *ðaalika* and its antecedent in (7) is almost equal since the proximal anaphor immediately follows its antecedent in (6) while there is only one word (i.e. *wa* ‘and’) separating the distal anaphor from its antecedent in (7). To be more precise, one can notice here that there are no words between the proximal anaphor and its antecedent in (6) since *fa-* ‘so’ is a bound morpheme attached to the proximal *haaða*. In example (7), there is only one free morpheme which is the particle *wa* ‘and’ as an independent word between the distal anaphor *ðaalika* and its antecedent.¹⁹ As a result, accessibility of CA anaphors cannot be adequately or satisfactorily measured based on anaphoric distance considerations. This result reflects those obtained by Botley and McEnery (2001: 226–229) who raise doubts regarding the reliability of anaphoric distance as a measure of accessibility of anaphoric demonstratives in English. Referring to Ariel’s notion of anaphoric distance, Botley and McEnery also explain that there are a number of examples in their corpus that “present methodological difficulties not only with measuring distance but also in relying on it as a valid measure at all” (2001: 226).

It can also be added that the results of this study support the findings attained by Botley and McEnery (2001: 224) in that “Ariel’s claim that proximal demonstratives will refer to antecedents that are textually closer than is the case with distal demonstratives has not received strong support” (see also Reboul 1997: 91). Based on this, it could be argued that the choice of demonstratives in CA is not merely spatial. This is similar to what many studies find about the use of exophoric proximals and distals. For instance, in her analysis of demonstratives in Finnish, Laury (1997:145–46) believes that “the use of

¹⁹ According to the definition of a *word* that is adopted for the purposes of this paper (see section 3.2 above), *wa* ‘and’ is an independent morphologic/orthographic word since it is a *harf* (i.e. particle) that has a meaning in CA.

demonstratives is not based on actual, concrete distance, but rather speakers use them to call attention to particular referents and indicate accessibility” (cited in Marchello-Nizia 2005: 59; see also Hanks 1990).

In conclusion, although Ariel (1988, 1999) focuses mainly on anaphoric distance as a major determinant of accessibility concerning the use of referring expressions in English, it emerges that this criterion cannot be taken as a reliable factor concerning the use of anaphoric demonstratives in CA. The average anaphoric distance measurements for the CA proximal and distal masculine anaphors contradict those for the feminine anaphors in the investigated CA texts. Moreover, the average anaphoric distance for the masculine proximal anaphor *haaḍa* is longer than that for the masculine distal *ḍaalika*, which runs counter to Ariel’s (1988, 1999) claim that proximals are associated with antecedents at short distance in comparison to distals as associated with antecedents at longer distance in discourse. Distance measurements for the CA masculine anaphors, in particular, indicate that anaphoric distance cannot be taken as a major or consistent determinant of the choice of proximal and distal anaphors in CA.

5. Conclusion and suggestions for further study

Ariel’s (1988; 1990) claim that proximals are associated with antecedents at shorter anaphoric distance than that for distals could not be substantially supported by the results of anaphoric distance measurements of the CA anaphors examined in this study. Textual distance between anaphors and their antecedents in CA is not a reliable parameter when it comes to discovering the determinants of the anaphoric use of demonstratives.

Considering the observations made in the present article and their contributions to our overall understanding of the use of anaphoric demonstratives in CA, there are some points that need to be noted in a critical fashion. Firstly, the issue of how proximal and distal anaphors are used in CA deserves further analysis to reveal the factors and determinants related to how these anaphors are used in discourse. Secondly, I examined direct and indirect anaphora as representing the same category (i.e. both are endophoric) based on their shared feature that they function within texts rather than in deictic contexts. However, further studies can investigate the determinants of using direct in comparison to using indirect anaphora in CA. The purpose here would be to know if these different functions have similar or different conditions of use.

The third point relates to the issue that the basic function of many anaphors involves sharedness of referentiality since the anaphor and its antecedent usually point at the same referent (see section 2.1 above). The feature of sharedness has different manifestations in the literature on demonstratives; for instance, Strauss (2002:135) refers to “sharedness or presumed sharedness of information” as one of the factors that determine a speaker’s choice of demonstratives. Hanks (1990: 47) describes context of interaction as *symmetric* or *asymmetric* depending on whether interactants share or do not share access to the referent. Shared access “may derive from common background experience, a shared perceptual field, a shared focus of attention, or other symmetrical relations.” Hanks’ (1990) concepts of ‘symmetry’ and ‘asymmetry’ in addition to Strauss’s (2002) notions of ‘focus’ and ‘sharedness of information’ seem to offer appropriate and promising

‘tools’ that can be employed to examine the major factors determining the choice and use of proximal and distal anaphors in CA.

The idea of *sharedness* to describe the relation between a referential expression and its antecedent can be easily recognized in Ariel’s (1988; 1990; 2001) concept of *unity*. *Unity* in Ariel (2001: 29) refers to whether the anaphor and its antecedent are “within vs. without the same frame/ world/ point of view/ segment or paragraph.” However, although Ariel (1988; 1990; 2001) does not investigate *unity* in relation to demonstratives, it is expected that investigating sharedness of “the same frame/ world/ point of view/ segment or paragraph” (Ariel 2001: 29) between an anaphor, its antecedent, and their referent, can offer significant insights concerning how CA anaphors are used in discourse.

Another issue is that although Ariel (1990) investigates *saliency* mainly in relation to High Accessibility Markers rather than to Intermediate ones such as anaphoric demonstratives, this should not be taken to indicate that the effect of such a factor is irrelevant to the use of demonstratives. Perhaps further research could reveal if *saliency* has any effects on the accessibility degree associated with the use of anaphors in CA. Further studies making use of larger or different types of corpora could help in determining if *saliency* is a relevant factor of accessibility in CA.

It is hoped here that future investigation of the possible determinants of using anaphors in CA will illustrate the importance of corpus-based studies in testing, modifying, or simply expanding the dominant perspectives and theories on anaphora. Such further research could perhaps also assist in the general aim towards providing tools and methods for a comprehensive understanding of the cognitive nature of anaphora in language.

References

- Ariel, M. 1988. Referring and accessibility. *Journal of Linguistics*, 24: 65-87.
- Ariel, M. 1990. *Accessing Noun-phrase Antecedents*. London: Routledge.
- Ariel, M. 2001. Accessibility theory: An overview. In T. Sanders, J. Schliperoord and W. Spooren (eds) *Text Representation* (Human Cognitive Processing Series). Philadelphia and Amsterdam: John Benjamins: 29-87.
- Beeston, A. 1970. *The Arabic Language Today*. London: Hutchinson.
- Bentivoglio, P. 1983. Topic continuity and discontinuity in discourse: A study of Spoken Latin American Spanish. In T. Givón (ed) *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Philadelphia and Amsterdam: John Benjamins: 255-312.
- Botley, S. and T. McEnery (eds). 2000. Discourse anaphora: The need for synthesis. In S. Botley and T. McEnery (eds) *Corpus-based and Computational Approaches to Discourse Anaphora*. Philadelphia and Amsterdam: John Benjamins: 1-39.
- Botley, S. and T. McEnery. 2001. Proximal and distal demonstratives: A corpus-based study. *Journal of English Linguistics*, 29: 214-233.
- Botley, S. 2006. Indirect anaphora: Testing the limits of corpus-based linguistics. *International Journal of Corpus Linguistics*, 11: 73-112.

- Brockelmann, C. 1977. *Tariikh Aladab Al3arabi* [History of Arabic Literature], vol. 1, 4th. edn. Abdul Haliim Najjar (trans). Cairo: Dar Alma3aarif.
- Cornish, F. 1996. 'Antecedentless' anaphors: Deixis, anaphora, or what? Some evidence from English and French. *Journal of Linguistics*, 32: 19-41.
- Cornish, F. 2008. How indexicals function in texts: Discourse, text, and one neo-Gricean account of indexical reference. *Journal of Pragmatics*, 40: 997-1018.
- Diessel, H. 1999. *Demonstratives: Form, function and grammaticalization*. Amsterdam and Philadelphia: John Benjamins.
- Diessel, H. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, 17: 463-489.
- Diver, W. 1984. *The Grammar of Modern English* (Textbook Linguistics G6801). Unpublished work.
- Dixon, R. 2003. Demonstratives: Cross-linguistic typology. *Studies in Language*, 27: 61-112.
- Ehlich, K. 1982. Anaphora and deixis: Same, similar or different? In Robert Jarvella & Wolfgang Klein (eds.), *Speech, Place and Action*. Chichester: John Wiley: 315-338.
- Francis, G. 1994. Labelling discourse: An aspect of nominal-group lexical cohesion. In Malcolm Coulthard (ed.), *Advances in Written Text Analysis*. London: Routledge: 83-101.
- Gasser, M. 1983. Topic continuity in written Amharic narrative. In Talmy Givón (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam and Philadelphia: John Benjamins: 95-140.
- Givón, T. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam and Philadelphia: John Benjamins: 1-41.
- Grosz, B. J, A. Joshi and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21: 203-225.
- Grundy, P. 2000. *Doing Pragmatics* (2nd edn). New York: Oxford University Press.
- Halliday, M. A. K. and R. Hassan. 1976. *Cohesion in English*. London: Longman.
- Hanks, W. 1990. *Referential Practice: Language and Lived Space among the Maya*. Chicago: The University of Chicago Press.
- Hasselbach, R. 2007. Demonstratives in Semitic. *Journal of the American Oriental Society*, 107: 1-27.
- Himmelmann, N. P. 1996. Demonstratives in narrative discourse: A taxonomy of universal uses. In B. Fox (ed) *Studies in Anaphora*. Amsterdam and Philadelphia: John Benjamins: 205-54.
- Hinds, J. 1983. Topic continuity in Japanese. In T. Givón (ed) *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins: 43-93.
- Ibn 3aqiil, A. 1996. *SharH Ibn 3aqiil 3ala Alfiat Ibn Maalik* [Ibn 3aqiil's explanation of Ibn Maalik's Alfia poem], vol 1. M. Halaawi (ed). Beirut: dar ?Hiia? Alturaath Al3arabi.
- Jaggar, P. 1983. Some dimensions of topic-NP continuity in Hausa narrative. In T. Givón (ed) *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam and Philadelphia: John Benjamins: 365-424

- Jarbou, S. O. 2010. Accessibility vs. physical proximity: An analysis of exophoric demonstrative practice in Spoken Jordanian Arabic. *Journal of Pragmatics*: 42, 3078-3097.
- Jarbou, S. O. 2012. Medial deictic demonstratives in Arabic: Fact or fallacy. *Pragmatics*, 22: 103-118.
- Laury, R. 1997. *Demonstratives in Interaction: The Emergence of a Definite Article in Finnish*. Amsterdam and Philadelphia: John Benjamins.
- Levinson, S. C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, S. C. 2006. Deixis and pragmatics. In L. Horn and G. Ward (eds) *The Handbook of Pragmatics*. Malden, MA: Blackwell Publishing: 97–121.
- Lyons, J. 1978. *Semantics*. Cambridge: Cambridge University Press.
- Marchello-Nizia, C. 2005. Deixis and subjectivity: The semantics of demonstratives in Old French (9th–12th Century). *Journal of Pragmatics*, 37: 43–68.
- Owens, J. 2006. *A Linguistic History of Arabic*. Oxford: Oxford University Press.
- Plag, I. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
- Reboul, A. 1997. What (if anything) is accessibility?: A relevance-oriented criticism of Ariel's accessibility theory of referring expressions. In J. Connolly, R. Vismans, C. Butler and R. Gatward (eds) *Discourse and pragmatics in functional grammar*. Berlin: de Gruyter: 91-108.
- Safwat, A. (comp) .1933. *Jamaharat Khotab Al-Arab* 'Collection of the public speeches/sermons of the Arabs', vol. 1. Beirut: Almaktaba Al3ilmiya.
- Sidner, C. 1983. Focusing in the comprehension of definite anaphora. In M. Brady and R. Berwick (eds) *Computational Models of Discourse*. Cambridge, MA: MIT Press: 267-330.
- Seuren, P. 2009. Primary and donkey anaphora. In P. Seuren (ed) *The Logic of Language: Language from within*, vol.2. Oxford: Oxford University Press: 283-310.
- Strauss, S. 2002. This, that, and it in Spoken American English: A demonstrative system of gradient focus. *Language Sciences* 24: 131-152.
- Taboada, M. 2008. Reference, centers, and transitions in Spoken Spanish. In J. K. Gundel and N. Hedberg (eds) *Reference: Interdisciplinary Perspectives*. Oxford: Oxford University Press: 167-215.
- Taboada, M. and L. Zabala. 2008. Deciding on units of analysis within Centering Theory. *Corpus Linguistics and Linguistic Theory*, 4: 63-108.
- Versteegh, K. 2001. *The Arabic Language*. Edinburgh: Edinburgh University Press.
- Walker, M. A, A. K. Joshi and E. F. Prince (eds). 1998. *Centering Theory in Discourse*. Oxford: Clarendon.

About the authors

Samir Jarbou obtained his PhD in rhetoric and linguistics from Indiana University of Pennsylvania, USA in 2002. He is currently employed at the Department of English Language and Linguistics, Jordan University of Science and Technology. He is mainly interested in deixis, anaphora, speech acts, and the semantics-pragmatics interface.

Fathi Migdadi is employed at the Department of English Language and Linguistics, Jordan University of Science and Technology. He received his PhD from Ball State

University, USA in 2003. His research interests include speech acts, conversation analysis, and bilingualism.

Appendix A

Following is a list of the symbols and what they stand for as used in this paper:

Symbol	Description
/ʔ/	glottal stop
/ð/	voiced interdental fricative
/ʕ/	voiced pharyngeal fricative
/H/	voiceless pharyngeal fricative
/kh/	voiceless velar fricative
/T/	emphatic voiceless alveolar
/dh/	emphatic voiced interdental fricative
/gh/	voiced velar fricative
/Q/	voiceless uvular stop
/S/	emphatic voiceless alveolar fricative
/dʒ/	voiced palato-alveolar affricate
/a/-/aa/, /i/-/ii/, /u/-/uu/	short-long vowels
Sg.	singular
Dl.	dual
Pl.	plural
F.	feminine
M.	masculine

Appendix B

The following table illustrates the majority of singular, dual, and plural demonstratives in CA; some of these are more common than others - the very rare ones have not been included.

Demonstrative	Description	Meaning
<i>haaða</i>	Sg. M.	‘this’
<i>ðaʔ</i>	Sg. M.	‘this’
<i>haaðihi, ðih, ðii, tihi, ðihi, tii, ðaat, tih, taa</i>	Sg. F.	‘this’
<i>(haa)ðani</i>	Dl. M.	‘these two’
<i>(haa)tani</i>	Dl. F.	‘these two’
<i>(haa)ʔula</i>	Pl. M/F.	‘these’
<i>(haa)ʔulaaʔi</i>	Pl. M./F.	‘these’
<i>ðaaka</i>	Sg. M	‘that’
<i>ðaalika</i>	Sg. M.	‘that’
<i>tiika</i>	Sg. F.	‘that’
<i>taalika</i>	Sg. F.	‘that’
<i>ðaanika</i>	Dl. M.	‘those two’

<i>taanika</i>	DI. F.	‘those two’
<i>ḍaannika</i>	DI. M.	‘those two’
<i>taannika</i>	DI. F.	‘those two’
<i>ṛulaaka</i>	Pl. M/F.	‘those’
<i>ṛulaaṛika</i>	Pl. M/F.	‘those’